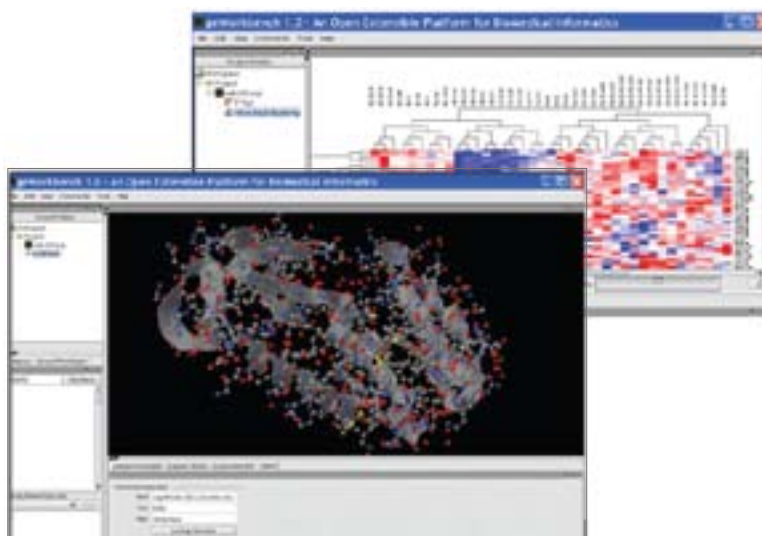


geWorkbench



Microarray gene expression and sequence data management

geWorkbench is an innovative open-source software platform for integrative genomic data analysis that provides scientists with transparent access to a number of external data sources through caGrid as well as to algorithmic services. geWorkbench also provides users with access to more than 50 built-in tools for data analysis and visualization and provides support for pathways (BioCarta), gene ontologies, networks and patterns based on regular expressions (i.e., pattern matching).



geWorkbench interface

Capabilities

- Load data from local or remote data sources and visualize gene expression, molecular interaction network, protein sequence and protein structure data in a variety of ways
- Access client- and server-side computational analysis tools such as t-test analysis, hierarchical clustering, self-organizing maps, regulatory networks reconstruction, BLAST searches and pattern/motif discovery
- Validate computational hypotheses through the integration of gene and pathway annotation information from curated sources as well as through Gene Ontology enrichment analysis
- Access analysis and visualization tools for microarray-based gene expression profiling data from a variety of systems, including Affymetrix MAS5/GCOS; Matrix format (geWorkbench); tab-delimited simple spreadsheet (GEO series matrix, RMAExpress) and GenePix; as well as for gene and protein sequence data (FASTA)
- Utilize the program's capacity to provide a wide range of plug-in components with analytical capabilities such as filter and normalize, promoter analysis, regulatory networks, differential expression, enrichment analysis, annotation, sequence analysis and pattern discovery

Categories of Use

- | | | | |
|--|--|--|---|
| <input type="checkbox"/> Biospecimens | <input type="checkbox"/> Data Sharing | <input type="checkbox"/> Imaging | <input type="checkbox"/> Proteomics |
| <input type="checkbox"/> Clinical Trials Management | <input type="checkbox"/> Genome Annotation | <input checked="" type="checkbox"/> Microarrays | <input type="checkbox"/> Translational Research |
| <input checked="" type="checkbox"/> Data Analysis & Statistical Tools | <input type="checkbox"/> Infrastructure | <input checked="" type="checkbox"/> Pathways | <input type="checkbox"/> Vocabularies |



- Access sophisticated tools from the MAGNet Center (<http://magnet.c2b2.columbia.edu>) for the analysis and reverse engineering of regulatory networks and for the integration of information from protein structures
- Achieve integrated access to many external data sources and computational services (e.g., GoldenPath at Santa Cruz, NCBI BLAST, BioCarta diagrams through caBIO)

Architecture Overview

- **Application type:** geWorkbench is a desktop application that also provides access to remote data and analysis services.
- **System requirements:** geWorkbench is downloaded as a self-installing package with support for Windows (XP and Vista), Linux and MAC OS-X; it requires the installation of Sun Java J2SE (Java JRE included with installer versions). A generic platform-independent version is also available.

Installation and Administration

- **Skill sets needed:** Users should have basic computer skills as well as familiarity with basic biology and the goals of the data analysis to be performed.
- **Infrastructure needed:** Installation of Java JRE is required before installing geWorkbench; a Windows version of geWorkbench can be downloaded with its own local Java JRE included.
- **Long-term administration needs:** None is required; however, some modules maintain local copies of genomic or sequence information that users may wish to update periodically (or wait for the next geWorkbench release).

Resources

| | |
|---|---|
| Tool Overview Page | https://cabig.nci.nih.gov/tools/geWorkbench |
| Primary Workspace | Integrative Cancer Research (ICR) https://cabig.nci.nih.gov/workspaces/ICR |
| Molecular Analysis Tools Knowledge Center | https://cabig-kc.nci.nih.gov/Molecular/KC/index.php/Main_Page |
| caBIG® Tool Inventory | https://cabig.nci.nih.gov/inventory |
| NCI Center for Bioinformatics Applications Support | ncicb@pop.nci.nih.gov |

Key Contributors:

- Center for Computational Biology and Bioinformatics at Columbia University
- Jackson Labs
- Northwestern University
- Washington University

Other Life Sciences Distribution Components:

- caArray
- Cancer Genome-Wide Association Studies (caGWAS)
- caTissue Core
- Clinical Trials Object Data System (CTODS)
- National Biomedical Imaging Archive (NBIA)



NIH Publication No. 09-7447
Printed September 2009

